

Рецензия

на проф. д-р Калинка Михайлова Калоянова - ФМИ, СУ "Св. Кл. Охридски"
за дисертационен труд за придобиване на образователната и научна степен "доктор"
на тема "Оптимизиране на стратегиите на заявки в бази данни с фиксирано
вертикално разделяне и в дистрибутирани база данни и тяхното
приложение в бази данни на семантичния интернет"
с автор Йенс Колер

Със заповед N 205 /1.12.2017г. на директора на Института по информационни и комуникационни технологии - БАН съм назначена за член на научно жури във връзка с придобиване на ОНС "доктор" в докторска програма "Информатика" (Научна област: 4. Природни науки, математика и информатика; Професионално направление: 4.6. Информатика и компютърни науки) от Йенс Колер, представил дисертационен труд на тема "Оптимизиране на стратегиите на заявки в бази данни с фиксирано вертикално разделяне и в дистрибутирани база данни и тяхното приложение в бази данни на семантичния интернет".

За рецензиране ми бяха предоставени два документа на английски език:

- дисертационен труд на тема "Оптимизиране на стратегиите на заявки в бази данни с фиксирано вертикално разделяне и в дистрибутирани база данни и тяхното приложение в бази данни на семантичния интернет";
- автореферат към дисертационния труд.

Обща информация за докторанта

Йенс Колер (Jens Kohler) завършва бакалавърска и магистърска степен във Faculty of Business and Computer Science, University of Applied Sciences, Albstadt-Sigmaringen (Germany). Работи като изследовател и лектор в няколко института в Германия в периода 2008-2014 година. От началото на 2015 година е лектор по Управление на знанията в Cooperative State University Mannheim (Duale Hochschule Baden-Wuerttemberg Mannheim). През 2015 година със заповед № 129-1/30.09.2015 г. е зачислен като докторант на самостоятелна подготовка за срок 01.10.2015 – 01.10.2017г. към Института по информационни и комуникационни технологии (ИИКТ) при БАН.

Анализ на съдържанието, резултатите и приносите на дисертационния труд

Изследванията на докторанта са свързани с тематиката за съхранение и достъп до информацията, като наред с традиционните въпроси, свързани с производителността на операциите за достъп до данните, се засягат и важните в последните години тенденции в областта - нарастване на обема на данните, сигурността на данните, представянето на данните в облачни среди.

Дисертационният труд включва увод, 8 глави, заключение, използвана литература, списък на използваните съкращения и няколко приложения, разпределени на 212 страници. Едно от приложенията илюстрира елементи от интерфейса на платформата SeDiCo, разработена от докторанта като среда за използване и оценяване на резултатите от изследванията. Останалите приложения предлагат списъци на фигури, таблиците и на елементи код, използвани в текста.

Уводната секция засяга мотивацията на работата на докторанта, свързана с търсене на възможности за оптимизация на използването на релационния модел, едновременно с това ориентирано и към повишаване на сигурността на данните, на базата на разпределено решение при използване на фиксирано вертикално разпределение на дялове при физическата организация на базата от данни. Тук е поставена целта на дисертационния труд, формулирани са няколко работни хипотези и произтичащите от тях задачи пред докторанта.

Глава първа представя основните концепции, които се използват в дисертационния труд. Дефинирани са редица концепции, като на места се навлиза в излишна детайлизация по отношение на представянето на добре познати понятия - например, тези от релационния модел и релационната алгебра. По-подробно са обяснени задачата за вертикално разделяне на релация и намирането на адекватни стратегии за опимизация на заявките към такава релация, както и работните хипотези, свързани с производителността на заявките.

Глава втора е посветена на въпросите, свързани с вертикалното разделяне на дялове при физическото съхранение на данните в релационни бази от данни (ДБ) и последващите особености при изпълнение на заявките върху такива релации. Теоретично е обоснована коректността на предложения метод при поддържане на цялостта на данните.

Глава трета се фокусира върху архитектурните концепции, залегнали в рамката SeDiCo. Засегнати са въпроси, свързани със сигурността и поверителността на данните,

представянето на данните в облачна среда, обектно-релационно съответствие, кеширане. Посочени са някои конкретни решения, които са имплементирани от докторанта - например, някои техники за следване на принципите за сигурност, цялостност и наличност на данните, както и за запазване на тяхната поверителност. Независимо от важността на изброените концепции, тяхното (на места твърде подробно) описание затруднява проследяването на основните резултати на дисертационния труд.

Глава четвърта представя на концептуално ниво използваната стратегия за оптимизация на заявките. Предложени са няколко подхода за намаляване на времето за отговор на заявките – пренаписване на заявките при отчитане на делението на дялове, кеширане, както и използване на SSD дискове вместо традиционните твърди дискове за съхранение на данните.

Конкретната реализация на посочените по-горе подходи е описана накратко в глава пета, а оценката на тяхното изпълнение е направена в шеста глава. Създадена е подходяща тестова среда и са проведени експерименти. Направени са и някои заключения от проведените експерименти.

Глава 7 обобщава получените основни резултати и ги съпоставя на работните хипотези, формулирани в първите глави, отчитайки кои хипотези са потвърдени и кои не са напълно потвърдени.

Глава 8 показва приложение на рамката SeDiCo при семантични Web приложения (RDF-базирани данни). И тук са разгледани въпросите, свързани с разпределението на данните, тяхната сигурност и производителността на заявките (в случая - на SPARQL). Реализацията е оценена и са направени изводи за използваемостта на подхода.

В заключението са обобщени приносите на направеното изследване. Тук са посочени явно публикациите, направени по тематиката на дисертационния труд. За всяка публикация е посочено към коя част от изследванията се отнася.

Приемам приносите, формулирани от автора, които могат да бъдат определени като научно-приложни:

- Дефиниране на принцип за разделяне на физическо ниво в релационна БД на основата на сигурността на данните;
- Разработване на механизъм за достъп (заявки) до вертикално разпределените данни;

- Реализиране на механизма за заявки и интегрирането му в разработената платформа SeDiCo;
- Провеждане на експерименти за оценяване на ефективността на предложените подходи;
- Трансфериране на предложените механизми в бази от данни на семантичния интернет.

Към посоченото по-горе бих добавила и комплексността на предложените решения, насочени към търсене на оптимизация, която засяга много направления - сигурност, производителност, използване на различни хардуерни решения и т.н.

Като цяло стилът на изложението в дисертационния труд е добър. Направени са редица теоретични обосновки на предложените решения. Предвид широтата на засегнатите проблеми, структурирането на съдържанието на дисертационния труд няма еднозначно решение. Все пак то би могло да бъде направено по-добре с оглед фокусиране върху представянето на резултатите.

Публикации, отразяващи дисертационния труд

Докторантът е направил 16 публикации, свързани с темата на дисертационния труд. Те са представени в периода 2012 - 2016 година главно в международни конференции, проведени в различни държави:

- 13 публикации на международни конференции;
- една публикация в списание "International Journal of Adaptive, Resilient and Autonomic Systems" (IJARAS);
- една публикация в сборник на издателство Springer Verlag - "Technologien fuer digitale Innovationen";
- една публикация в електронно списание eJournal of AKWI - "Anwendungen und Konzepte der Wirtschaftsinformatik" (на немски език).

Една от посочените публикации е самостоятелна, останалите са в съавторство.

Представени са данни за две забележани цитирания на публикации, в които участва докторантът - в дипломна работа на студент-магистър от австралийски университет и в публикация от международна конференция.

Автореферат

Авторефератът следва структурата на дисертационния труд. Представени са всички глави от дисертационния труд, като акцентът е върху резултатите.

Авторефератът отразява сравнително адекватно същността на изследванията, макар че това би могло да бъде постигнато и в по-малък обем от 59 страници.

Коментари, критични бележки и препоръки

Дисертационният труд на Йенс Колер представя работа по широк кръг проблеми, свързани с ефективното съхранение и достъп до данните. Положително е усилието на докторанта да търси интегрирани решения по тези въпроси, както и да предоставя теоретично обосновка на предложените решения. Извършен е голям обем от работа, както по отношение на изследователската част, така и като практическа имплементация на предложените решения.

За съжаление при представяне на различните концепции в почти всяка глава са направени множество обяснения и препратки към други източници, което разводнява презентацията и създава затруднения при определянето на действителните приноси на дисертацията.

Като цяло докторантът прилага широк набор от методи, техники и средства, за да получи крайните резултати, които показват добри познания в областта.

Добро впечатление прави големият брой дипломни работи, разработени под ръководството на Йенс Колер. Разпределянето на работата и съгласуваността на отделните теми изисква много усилия и свидетелства за добри умения за дефиниране на специфични теми и определяне на постъпково изпълнение на задачите, необходими за реализацията им.

Заклучение

В заключение считам, че дисертационният труд, представен от докторанта Йенс Колер, напълно покрива съвкупността от изисквания на ЗРАСРБ, ПЗРАСРБ и Правилника за специфичните условия за придобиване на научни степени и за заемане на академични длъжности в Института по информационни и комуникационни технологии при БАН и убедено предлагам на уважаемото научно жури да присъди на Йенс Колер образователната и научна степен "доктор" в професионално направление 4.6. "Информатика и компютърни науки".

26.02.2018 г.

София

